



# PROTEUS

**Scalable online machine learning for predictive analytics and real-time interactive visualization**

**687691**

---

## **D6.6 Report on scientific dissemination activities – V2**

---

**Lead Author: Hamid Bouchachia**

**Reviewers: Rachel Finn**

Deliverable nature:	Report (R)
Dissemination level: (Confidentiality)	Public (PU)
Contractual delivery date:	November 30 <sup>th</sup> 2018
Actual delivery date:	November 30 <sup>th</sup> 2018
Version:	3
Total number of pages:	17
Keywords:	Scientific dissemination

*Abstract*

The present deliverable describes the scientific dissemination activities and materials along with the time line and success indicators. It includes an updated record of activities related to scientific dissemination that have been undertaken during the first and second halves of the project. It reflects on the publications, software and events at which the consortium members took part.

---

## **Executive summary**

In the second reporting period (RP2) of PROTEUS, various scientific dissemination actions have taken place. They are directed to different audiences from the specialised public to specific public in order to illustrate the ambitions of the project and to enhance its global impact. The effort of promoting the scientific and applicative aspects of PROTEUS has continued during RP2.

A number of research outputs have resulted and disseminated to different audiences such as academia, general public and the business community. This report describes the dissemination activities undertaken in the second period of PROTEUS. It mainly presents and discusses the activities according to the following classification:

- 1) Dissemination events
- 2) Scientific publications
- 3) Software components

## Document Information

<b>IST Project Number</b>	687691	<b>Acronym</b>	PROTEUS
<b>Full Title</b>	Scalable online machine learning for predictive analytics and real-time interactive visualization		
<b>Project URL</b>	<a href="http://www.proteus-bigdata.com/">http://www.proteus-bigdata.com/</a>		
<b>EU Project Officer</b>	Martina EYDNER		

<b>Deliverable</b>	<b>Number</b>	D6.6	<b>Title</b>	Report on scientific dissemination activities – V2
<b>Work Package</b>	<b>Number</b>	WP6	<b>Title</b>	

<b>Date of Delivery</b>	<b>Contractual</b>	M36	<b>Actual</b>	M36
<b>Status</b>	version 3.0		final <input type="checkbox"/>	
<b>Nature</b>	report <input checked="" type="checkbox"/> demonstrator <input type="checkbox"/> other <input type="checkbox"/>			
<b>Dissemination level</b>	public <input checked="" type="checkbox"/> restricted <input type="checkbox"/>			

<b>Authors (Partner)</b>	BU			
<b>Responsible Author</b>	<b>Name</b>	Hamid Bouchachia	<b>E-mail</b>	<a href="mailto:abouchachia@bournemouth.ac.uk">abouchachia@bournemouth.ac.uk</a>
	<b>Partner</b>	BU	<b>Phone</b>	+44 1202 96 24 01

<b>Abstract (for dissemination)</b>	The present deliverable describes the scientific dissemination activities and materials along with the time line and success indicators. It includes an updated record of activities related to scientific dissemination that have been undertaken during the first and second halves of the project. It reflects on the publications, software and events at which the consortium members took part.
<b>Keywords</b>	Scientific dissemination, scientific events, scientific publications.

Version Log			
Issue Date	Rev. No.	Author	Change
September 30 <sup>th</sup> 2018	V.1	Hamid Bouchachia	Structure of the document
November 07 <sup>th</sup> , 2018	V.2	Hamid Bouchachia	First draft
November 25 <sup>th</sup> , 2018	V.2	Rachel Finn	Review comments
November 30 <sup>th</sup> , 2018	V.3	Hamid Bouchachia	Final version

## Table of Contents

Executive summary.....	3
Document Information.....	4
Table of Contents.....	5
List of Tables.....	6
1. Introduction.....	7
2. Dissemination Events and Publications.....	8
2.1. Dissemination events.....	8
2.2. Publications.....	10
3. Software.....	15
4. Conclusion.....	17

## List of Tables

Table 1: Dissemination activities.....	8
Table 2: Events attended (updated) .....	8
Table 3: List of publications.....	11
Figure 4: Papers submitted.....	14

## 1. Introduction

To enhance the visibility of PROTEUS, various dissemination activities have taken place during the second period of the project. In particular, the scientific dissemination under its different forms has been continuous taking a number of guidelines into account:

- Background Intellectual Property
- Foreground Intellectual Property
- Intellectual Property Rights of the partner (patents, copyright, trademark, etc.)
- Confidentiality of data (emanating mostly from industrial partners)
- Target prominent journals and conferences when publishing
- Open access whenever possible
- Rigorous referencing in all publications
- Diversification of the audiences
- Acknowledgment of the EC financial support

This document provides the details of the scientific dissemination activities undertaken in the context of PROTEUS. These include mainly publications, scientific event organization and attendance, as well as software developed during the project. The goal of the document is to highlight the ongoing effort of PROTEUS to present the outcome and enhance the awareness of various communities about the objectives and results of PROTEUS.

The remaining part of this document dedicated to scientific dissemination only consists of 2 sections. In the first of them, the dissemination events and publications that PROTEUS has produced so far are presented. Section 3 describes the software components developed. Section 4 summarises the report.

## 2. Dissemination Events and Publications

The consortium has been actively targeting a number of venues for disseminating the work so far achieved. In particular, various types of outcomes have been produced as shown in Table 1 below. These outputs have been dedicated to different audiences such as academia, general public, media, policy makers, and the business community.

**Table 1: Dissemination activities**

	Academia	General public	Media	EU policy makers	Business
Scientific Publication	✓	✓			✓
Talk	✓	✓		✓	✓
Poster	✓				✓
Conferences	✓	✓		✓	✓
Social media			✓		

In the following, the details of the dissemination activities are presented.

### 2.1. Dissemination events

Members of the PROTEUS consortium took part at many international scientific events to disseminate the project objectives, impact and scientific results to a wide audience. The following table summarises the events where PROTEUS was presented.

**Table 2: Events attended (updated)**

Event	Date	Type	Audience	Coverage
<b>RP1</b>				
The ACM SIGMOD Conference	26/06- 01/07/2016	Scientific	1000	International
Beyond MR Workshop co-located with the ACM SIGMOD Conference	26/06- 01/07/2016	Scientific	50	International
FiCloud Conference	23/08/2016	Scientific	50	International
Predictive APIs Conference	10-12/10/ 2016	Business	500	International
Big Data Spain	17-18/11/ 2016	Scientific / Business	1000	International
EuroPro Workshop in EDBT/ICDT	21-24/03/2017	Academic	200	International
Workshop STREAMEVOLV co-located with the European Conference on Machine Learning	19-23/09/2016	Scientific	400	International
Workshop on "Machine Learning Algorithms and Applications" co-located with the International Conference on Machine Learning and Applications	18-20/12/2016	Scientific	300	International
Big Data Coruña Summit	25-26/05/2017	Scientific / Business	50	International



Big Data Value Association Summit	1-2/12/ 2016	Scientific / industry / policy makers	450	European
<b>RP2</b>				
European Big Data Standards Workshop, Dublin 2017	14 August 2017	Scientific / industry / policy makers	50	European
The Growing Ubiquity of Algorithms in Society, The Royal Society, London 2017	13 September 2017	Scientific / industry / policy makers	50	UK
Workshop at the European Big Data Value Forum: Industrial Data Platforms for the Manufacturing domain	21st to 23rd November 2017	Scientific / industry / policy makers	500	European
LibreCon 2017, the Business and Open Technologies Conference	October 19-20, 2017	Scientific / industry	300	European
Flink Forward Berlin 2017	12 September 2017	Scientific / industry	100	International
The future of data analytics in the UK: growth, challenges and the policy context	7 December 2017	Industry / Policy	100	UK
AI Congress 2018, London	30-31 January 2018	Industry / Scientific	1000	International
SIRIUS Digital Twins workshop, Oslo	20 March 2018	Scientific / industry	30	European
Workshop on Software Architecture Challenges in Big Data	25 September, 2018	Scientific / industry	25	European
Smart Factories Workshop	26 September, 2018	Scientific / industry	80	International
18th UK Workshop on Computational Intelligence	September 5-7, 2018	Scientific	80	International
European Big Data Value Forum, 2018 Data driven business model workshop	November 13-15	Scientific / industry	500	European

In addition to presentations at academic conferences (please see Table 3 below), the members of the consortium have delivered public lectures about PROTEUS in business/academia meetings as shown in Table 2 above. In particular, and to name few, the project members co-organized again the *STREAMEVOLV* workshop co-located with the *European Conference on Machine Learning* in 2017 and 2018 respectively, as well as the *Workshop at the European Big Data Value Forum: Industrial Data Platforms for the Manufacturing domain* in 2017 and the *18th UK Workshop on Computational Intelligence* in 2018. So overall, we could execute all initially planned dissemination actions (2 workshops and pan-European conference) (please, see Section 2.2.5 of the proposal - Communication activities and tools, please see also Section Conclusion below).

## 2.2. Publications

PROTEUS is engaging in academic dissemination of results through high-impact journals and conferences. So top dissemination venues have been targeted like VLDB, Journal of Machine Learning, Journal of Neural Networks, IEEE Transaction on Neural Networks and Learning Systems, the European Conference on Machine Learning, and the IEEE International Conference on Data Engineering (ICDE), etc.

The consortium members have submitted research work (under review) in very prominent venues such as the IEEE transactions on Neural Networks and Learning Systems, Journal of Machine Learning Research, Neural Networks, Pattern Recognition, while some other publications are under preparation for others conferences in the areas of Machine Learning and Big Data (Please see Table 4).

Note also the funding source (EC) has been acknowledged in the publications, including those already published and those being under review or under preparation.

In terms of open access, all partners are aware about the importance of making the outcome of PROTEUS accessible to the general readers. PROTEUS as a consortium as well as the individual partners are primarily pursuing green open access. Since gold open access is not free, PROTEUS partners are asked to seek institutional contribution to cover the open access fees. Presently, most of the publications are made accessible, when they are not gold open access. They are provided in the form of “accepted version for publication” or “published version” after the embargo date (e.g., 6 months from the publication date) to avoid any copyright issues.

Moreover, many of the publications are joint private-public (PP) publications. Some of them are still under review. The consortium will be submitting all publications to EC via SyGMA.

Table 3 summarises the list of publications produced so far, while Table 4 covers publications submitted (under review). Others are still under preparation. This list exceeds the expectations by far as we initially targeted 8 publications and 8 attendances of conferences.

**Table 3: List of publications**

No	WP	DOI	Type	Repository Link	Title	Authors	Title of the Journal/Proceedings/Books series	Conf date Journal/Book details	Pages	ISBN/ISSN/eISSN	Publisher	Open access/made available	Peer Rev.	Joint PP pub.	Sub. to EC
<b>RP1</b>															
1	WP3	10.1145/2926534.2926540	Conf. paper	<a href="http://dl.acm.org/citation.cfm?id=2926540">http://dl.acm.org/citation.cfm?id=2926540</a>	Bridging the gap: towards optimization across linear and relational algebra	A. Kunft, A. Alexandrov, A. Katsifodimos, V. Markl	The 3rd ACM SIGMOD Workshop on Algorithms and Systems for MapReduce and Beyond	26/06/16	-	ISBN: 978-1-4503-4311-4	ACM	YES	YES	YES	YES
2	WP3	10.1145/2882903.2899396	Conf. paper	<a href="http://dl.acm.org/citation.cfm?id=2882903.2899396">http://dl.acm.org/citation.cfm?id=2882903.2899396</a>	Emma in Action: Declarative Dataflows for Scalable Data Analysis	A. Alexandrov, A. Salzmann, G. Krastev, A. Katsifodimos, V. Markl	ACM SIGMOD International Conference on Management of Data	26/06/16	2073-2076	ISBN: 978-1-4503-3531-7	ACM	NO	YES	YES	YES
3	WP3	-	Magazine paper	<a href="http://sites.computer.org/debull/A15dec/p28.pdf">http://sites.computer.org/debull/A15dec/p28.pdf</a>	Apache Flink: Stream and Batch Processing in a Single Engine	P. Carbone, S. Ewen, S. Haridi, A. Katsifodimos, V. Markl, K. Tzoumas	Bulletin of the IEEE Computer Society Technical Committee on Data Engineering	Vol. 38 No. 4, Issue on Next-Generation Stream Processing Systems, 12/ 2015	28-38	-	IEEE	YES	YES	NO	YES
4	WP3	10.1145/2949741.2949754	Journal paper	<a href="http://dl.acm.org/citation.cfm?doid=2949741.2949754">http://dl.acm.org/citation.cfm?doid=2949741.2949754</a>	Implicit Parallelism through Deep Language Embedding	A. Alexandrov, A. Katsifodimos, G. Krastev, V. Markl	ACM SIGMOD Record	Vol. 45, No .1 03/2016	51-58	ISSN:0163-5808	ACM	YES	YES	NO	YES
5	WP5	10.1109/W-FiCloud.2016.46	Conf. paper	<a href="http://ieeexplore.ieee.org/abstract/document/7592720/">http://ieeexplore.ieee.org/abstract/document/7592720/</a>	An incremental approach for real-time Big Data visual analytics	I. G. Fernández, R. C. Tejedor, A. Bouchachia	The IEEE 4th International Conference on Future Internet of Things and Cloud	22-24/08/16	177-182	ISBN: 978-1-5090-3946-3	IEEE	YES	YES	YES	YES
6	WP5	10.1016/j.patrec.2016.08.015	Journal paper	<a href="http://www.sciencedirect.com/science/article/pii/S0167865516302173">http://www.sciencedirect.com/science/article/pii/S0167865516302173</a>	Improving the efficiency of IRWLS SVMs using parallel Cholesky factorization	R. D. Morales, Á. N. Vázquez,	Pattern Recognition Letters	Volume 84, 1 December 2016	91-98	0167-8655	ELSEVIER	NO	YES		YES

7	WP4	10.1109/TNNLS.2016.2614393	Journal paper	<a href="http://ieeexplore.ieee.org/document/7605500/">http://ieeexplore.ieee.org/document/7605500/</a>	A Bi-Criteria Active Learning Algorithm for Dynamic Data Streams	S. Mohamad, A. Bouchachia, M.S. Mouchaweh	IEEE Transactions on Neural Networks and Learning Systems	N/A (early access)	1-13	ISSN:2162-2388	IEEE	YES	YES	NO	YES
8	WP4	N/A	Conf/Work Paper	<a href="http://eprints.bournemouth.ac.uk/24798/">http://eprints.bournemouth.ac.uk/24798/</a>	Aggregation Algorithm Vs. Average for Time Series Prediction	W. Jamil, Y. Kalnishkan, A. Bouchachia	ECML/PKDD 2016 Workshop on Large-scale Learning from Data Streams in Evolving Environments	STREAMEVOLV-2016, 23 September 2016	69-82	-	Ecole des Mines, France	YES	YES	NO	YES
9	WP4	N/A	Conf/Work Paper	<a href="http://eprints.bournemouth.ac.uk/24798/">http://eprints.bournemouth.ac.uk/24798/</a>	Active Learning for Data Streams under Concept Drift and concept evolution	S. Mohamad, M.S. Mouchaweh A. Bouchachia	ECML/PKDD 2016 Workshop on Large-scale Learning from Data Streams in Evolving Environments	STREAMEVOLV-2016, 23 September 2016	51-68	-	Ecole des Mines, France	YES	YES	NO	YES
10	WP4	10.1109/EAIS.2015.7368803	Conf. paper	<a href="http://ieeexplore.ieee.org/document/7368803/?reload=true&amp;arnumber=7368803">http://ieeexplore.ieee.org/document/7368803/?reload=true&amp;arnumber=7368803</a>	A non-parametric hierarchical clustering model	S. Mohamad, A. Bouchachia, M.S. Mouchaweh	The 2015 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS)	01-03/12/2016	1-7	ISBN: 978-1-4673-6698-4	IEEE	YES	YES	NO	YES
11	WP4	10.1142/9789814675017_0005	Book chapter	<a href="http://www.worldscientific.com/worldscibooks/10.1142/9548">http://www.worldscientific.com/worldscibooks/10.1142/9548</a>	Fuzzy Classifiers	A. Bouchachia	Handbook on Computational Intelligence	01/05/2016		ISBN: 978-981-4675-00-0	World Scientific	YES	YES	NO	YES
12	WP4	10.1007/s00500-015-1946-4	Journal paper	<a href="http://link.springer.com/article/10.1007/s00500-015-1946-4">http://link.springer.com/article/10.1007/s00500-015-1946-4</a>	MSAFIS: an evolving fuzzy inference system.	J. J. Rubio, A. Bouchachia	Softcomputing	Vol. 21, No. 9 May 2017,	2357 - 2366	Online ISBN:1433-7479	Springer	YES	YES	NO	NO
<b>RP2</b>															
13	WP3	<a href="http://www.vldb.org/pvldb/vol10/p2061-kunft.pdf">10.14778/3151106.3151110</a>	Conf. paper	<a href="http://www.vldb.org/pvldb/vol10/p2061-kunft.pdf">http://www.vldb.org/pvldb/vol10/p2061-kunft.pdf</a>	Blockjoin: efficient matrix partitioning through joins	A. Kunft, Asterios Katsifodimos, Sebastian Schelter, Tilmann	Proceedings of the VLDB Endowment 10.13	Volume 10 Issue 13	2061-2072		ACM	YES	YES	NO	YES

						Rabl, Volker Markl									
14	WP3	10.1109/ ICDE.20 18.00135	Proce- ding s	<u>Scotty: Efficient Window Aggregation for out-of- order Stream Processing</u>		J. Traub, P. Grulich, A. R. Cuellar, S. Breß, A. Katsifodimos, T. Rabl, V. Markl	IEEE International Conference on Data Engineering (ICDE)	25 October 2018		2375- 026X	IEEE	NO	NO	NO	YES
15	WP3	10.1109/ ICDE.20 18.00169	Conf. paper	<a href="https://ieeexplore.ieee.org/document/8509390">https://ieeexpl ore.ieee.org/d ocument/8509 390</a>	Benchmarking Distributed Stream Data Processing Engines	Jeyhun Karimov ; Tilman Rabl ; Asterios Katsifodimos ; Roman Samarev ; Henri Heiskanen ; Volker Markl	IEEE 34th International Conference on Data Engineering (ICDE)	April 16-19, 2018	1507- 1518		IEEE	YES	YES	NO	YES
16	WP4	N/A	Book (Proc eedin gs)	<a href="https://link.springer.com/bookseries/11156">https://link.spr inger.com/boo kseries/11156</a>	Advances in Computational Intelligence Systems	A. Lotfi, A. Bouchachia, A. Gegov, C. Langensiepen, T. McGinnity	The 18th UK Workshop on Computational Intelligence	September 5-7, 2018	-	ISBN 978-3- 319- 97981-6	Springer	NO	YES	NO	NO
17	WP4	<a href="https://doi.org/10.1016/j.eswa.2017.10.026">https://do i.org/10. 1016/j.es wa.2017. 10.026</a>	Journ al paper	<a href="https://doi.org/10.1016/j.eswa.2017.10.026">https://doi.org/ 10.1016/j.esw a.2017.10.026</a>	Batch-based active learning: Application to social media data for crisis management.	D. Pohl, A. Bouchachia, H. Hellwagner	Expert Syst. Appl.	Vol. 93 March 2018	232- 244	ISSN 0957- 4174	Elsevier	NO	YES	NO	YES
18	WP4	<a href="https://doi.org/10.1016/j.neunet.2017.10.004">https://do i.org/10. 1016/j.ne unet.201 7.10.004</a>	Journ al paper	<a href="https://www.sciencedirect.com/science/article/pii/S0893608017302435">https://www.s ciencedirect.c om/science/art icle/pii/S0893 60801730243 5</a>	Active learning for classifying data streams with unknown number of classes.	S. Mohamad, M.S. Mouchaweh A. Bouchachia	Neural Networks	Vol. 98 February 2018	1-15	ISSN 0893- 6080	Elsevier	NO	YES	NO	YES
19	WP4	<a href="https://doi.org/10.1145/3241403.3241438">10.1145/ 3241403. 3241438</a>	Conf. paper	<a href="https://dl.acm.org/citation.cfm?id=3241438">https://dl.acm. org/citation.cf m?id=324143 8</a>	Scalable online learning for Flink: SOLMA library	W. Jamil N.-C. Duong, W. Wang, C. Mansouri, S. Mohamad, A. Bouchachia	The 12 <sup>th</sup> European Conference on Software Architecture	ECSA (Companion) September 24 - 28, 2018	1-4	ISBN: 978-1- 4503- 6483-6	ACM	NO	YES	NO	YES

20	WP4		Conf. paper		Model Selection in Online Learning for Times Series Forecasting	W. Jamil A. Bouchachia	The 18th UK Workshop on Computational Intelligence	September 5-7, 2018	83-95	ISBN 978-3-319-97981-6	Springer	NO	YES	NO	YES
21	WP4	N/A	Conf/Work Paper	<a href="http://dblp.org/db/conf/pkdd/iotstreaming2017.html">http://dblp.org/db/conf/pkdd/iotstreaming2017.html</a>	Workshop Proceedings 1958	M.S. Mouchaweh A. Bifet, A. Bouchachia, J. Gama, R. Ribeiro	ECML/PKDD 2017 Workshop on IoT Large Scale Learning from Data Streams	September 18-22, 2017		-	CEUR	YES	YES	NO	NO
22	WP4	N/A	Journal paper	N/A	Competitive Online Regularised Regression	W. Jamil A. Bouchachia	Machine Learning	Accepted with revision		ISSN 1573-0565	Springer	NO	YES	NO	YES

Table 4: Papers submitted

Work package	Type of output	Title	Venue
WP3	Workshop paper	PolyBench: The first benchmark for polystores	Tenth TPC Technology Conference on Performance Evaluation & Benchmarking (TPCTC 2018)
W4	Journal paper	Competitive Normalised Least Squares Regression	IEEE Transaction on Neural Networks and Learning Systems
W4	Journal paper	Online Weighted Averaging Passive Aggressive Learning	IEEE Transaction on Neural Networks and Learning Systems
W4	Conf. paper	Aggregating Algorithm for Iterated Ridge Regression	Algorithmic Learning Theory Conference
W4	Conf. paper	Online Bayesian Shrinkage Regression	European Conference on Neural Networks, Computational Intelligence and Machine Learning
W4	Conf. paper	Online Bi-level Stochastic Gradient for Support Vector Machines	European Conference on Neural Networks, Computational Intelligence and Machine Learning
W4	Conf. paper	Competitive Online Regularised Regression	European Conference on Machine Learning and Practice of Knowledge Discovery in Databases
W4	Conf. paper	Asynchronous Stochastic Variational Inference	European Conference on Neural Networks, Computational Intelligence and Machine Learning

### 3. Software

A number of software components have been developed in the context of PROTEUS. In the following a short description of each of them is presented. Please note that all software products/modules are open-source.

#### a) PROTEUS Engine

This is an overhauled version of Apache Flink supporting hybrid computation on batch datasets and data streams. Details about PROTEUS Engine can be found here: <https://github.com/proteus-h2020/proteus-engine>.

#### b) PROTEUS Language

A declarative language library, called *Emma*, has been developed to meet the need of Scalable Data Analysis. *Emma* aims at improving the developer productivity by hiding parallelism aspects behind a high-level, declarative API. *Emma* supports state-of-the-art dataflow engines like Apache Flink and Apache Spark as backend co-processors. More information about the language is available at <http://emma-language.org> and further details are located at: <https://github.com/proteus-h2020/proteus-language>.

#### c) PEACH (Proteus Elastic Cache)

PEACH is a distributed key-value cache that can be used both inside and outside of the PROTEUS scope. The cache aims to provide low latency responses on a distributed elastic deployment with fault-tolerance capabilities. As a generic design, the cache could be integrated within Apache Flink to speedup computing processes. More information can be found at: <https://github.com/proteus-h2020/peach>.

#### d) PROTEUS Incremental Analytics

A backend module that implements incremental version ( $\sim O(1)$  computational cost using approximations) of most common analytics operations. Proteus-backend is implemented on top of the Apache Flink streaming engine. Further details can be found at: <https://github.com/proteus-h2020/proteus-backend>.

#### e) SOLMA (Scalable Online Machine Learning and Data Mining Algorithms)

A scalable library adapted to the data analytics platform, Apache Flink. It consists of efficient distributed online algorithms for basic utilities, sketches as well as advanced online predictive analytics for tasks like classification, clustering, regression, ensemble methods, and novelty and change detection. Currently, the library encompasses a number of algorithms:

##### 1) Basic algorithms:

- Online moments: simple mean, simple variance, weighted mean, weighted variance, exponentially weighted mean and variance, moving average, aggregation algorithm.
- Online sampling: Simple reservoir sampling, weighted reservoir sampling and adaptive reservoir sampling

- Online frequent directions
  - Incremental principal component analysis
- 2) Classification algorithms
- Online support vector machines (OSVM)
  - Online bi-level stochastic gradient for support vector machines (OBSG-SVM)
  - Online passive-aggressive algorithms (PA)
- 3) Regression algorithms
- Online least absolute shrinkage and selection operator (LASSO)
  - Online ridge regression (ORR)
  - Online shrinkage via limit of Gibbs sampling (OSLOG)
  - Aggregating algorithm for regression (AAR)
  - Competitive online iterated ridge regression (COIRR)
- 4) Drift handling and anomaly detection
- Online weighted averaging passive-aggressive algorithm (WAPA)
  - Online normalised least mean square regression (ONLMSR)
  - Anomaly detection using incremental PCA
- 5) Topic modelling
- Latent Dirichlet allocation (LDA)

All these algorithms are readily available in SOLMA, some other algorithms like semi-supervised algorithms will be further implemented in SOLMA. It is very important to mention that SOLMA will be populated with other algorithms, especially with clustering algorithms as we intend to maintain SOLMA in the future. All algorithms are available on Github: <https://github.com/proteus-h2020/SOLMA>.

#### f) PROTEIC.JS

To accommodate visualization and visual analytics for big data, an HTML5 and CSS3 charts library has been developed. It is adapted to work for both data-at-rest and data-in-motion. More details can be found at: <https://github.com/proteus-h2020/proteic>.



## 4. Conclusion

The present document provides a short outline of the scientific dissemination activities undertaken in PROTEUS. In particular the list of events the consortium members participated in, a detailed list of publications produced and the list of software components developed in the context of PROTEUS.

The related KPI set in the proposal (Section 2.2.5) have been all met as indicated the table below:

<b>Communication activity</b>	<b>Objective</b>	<b>Met (Y/N)</b>
Peer reviewed papers	At least 3 high-impact JCR scientific publications	Y
Conference presentations	At least 10 participations in congresses	Y
Dedicated symposia, workshops, conference	Hosting 1 final workshop, 150 people expected. Attending 2 workshops (one with stakeholders and one with experts/related parties) to advertise the project. 1 pan-European conference at the end of the project to disseminate the findings and outputs.	Y
News release	At least 3 news release a year per participant country.	Y
Social media	ResearchGate; dedicated LinkedIn Group; Twitter	Y
Community engagement	Contributing the enhancements made by PROTEUS back to the Apache community, make these changes available to the public in order to gain attraction in the Apache Flink community.	Y
Newsletters	At least 1 newsletter a year	-

As a future work, we will continue to maintain PROTEUS software and pursue the dissemination work including the publication effort.